



Kai Berger,
Christian Lipski,
Christian Linz,
Anita Sellent,
Marcus Magnor

**A ghosting artifact detector for interpolated image quality
assessment**

Braunschweig : Computer Graphics Lab, 2009

**(Technical report / Computer Graphics Lab, TU
Braunschweig ; 2009-7-10)**

Veröffentlicht: 22.07.2009

<http://www.digibib.tu-bs.de/?docid=00028790>



KAI BERGER

berger@cg.tu-bs.de

Computer Graphics Lab, TU Braunschweig

CHRISTIAN LIPSKI, CHRISTIAN LINZ, ANITA SELLENT

{lipski,linz,sellent}@cg.tu-bs.de

Computer Graphics Lab, TU Braunschweig

MARCUS MAGNOR

magnor@cg.tu-bs.de

Computer Graphics Lab, TU Braunschweig

A ghosting artifact detector for interpolated image quality assessment

Technical Report 2009-7-10

July 21, 2009

Computer Graphics Lab, TU Braunschweig

Abstract

We present a no-reference image quality metric for image interpolation. The approach is capable of detecting ghosting artifacts, e.g., in image based rendering scenarios. Based on the assumption that ghosting artifacts can be detected locally, perceived visual quality can be predicted from the amount of regions that are affected by ghosting. Because the approach does not require any reference image, it is very suitable, e.g., for assessing quality of image-based rendering techniques in general settings.

1 Introduction

Free-viewpoint navigation around real-world dynamic scenes has recently received considerable attention in computer graphics as well as computer vision, either making use of reconstructed scene geometry [CTMS03, ZKU⁺04, VBK05, SH07, dAST⁺08] or being based on the input images alone [LH96, MP04, WSY07]. In both approaches, the quality of the rendered virtual image can safely be assumed to be lower than the input photos. Quality degradation in image interpolation typically manifests itself in image blurring or image ghosting artifacts, observers perceive as highly distracting.

In geometry-assisted systems, artifacts come from inexact reconstruction and/or camera calibration errors. Some of the visual degradation can be corrected by advanced reprojection techniques [EMD⁺08]. In purely image-based systems, the major sources of error are occlusions/disocclusions of different depth layers and inaccurate correspondences. The last one, in general, leads to prominent ghosting artifacts in the interpolated image. While those artifacts can be compensated for by correcting the correspondence fields [SLW⁺08b], this correction in the novel views remains a tedious and often subjective manual task.

The identification of various artifacts including blurriness, blockiness and noisiness has been extensively studied in the field of image/video compression and transmission and can be applied to the analysis of interpolated images as well [WS03, Mar04, FM05, SBC05, LH08]. However, only little research has been devoted to image quality metrics tailored to the particularities of free-viewpoint systems, i.e., detection of ghosting where no ground-truth data is available. In a first step in this direction, Stark and Kilner [SKH08, KSGH09] proposed a quality assessment system for free-viewpoint video production. Yet, the metrics used essentially measure the quality of 3D reconstruction in image space and are thus only applicable to geometry-based free-viewpoint systems.

As main contribution, this technical report presents a no-reference quality metric for purely image-based free-viewpoint systems, focusing on the detection of ghosting artifacts where no ground-truth reference images are available for comparison. Our approach is based on the observation that ghosting occurs when corresponding pixels from two or more original images I_1, I_2 are warped to different pixel positions in the interpolated image, Fig. 1. We validate the results of our ghosting detector against the results of a user study. Furthermore we tested the proposed ghosting metric against an image dataset obtaining only JPEG-compression artifacts and an image dataset obtaining only blurring artifacts.

The rest of this paper is organized as follows. After summarizing the related work in Sect. 2, we give a detailed description of our ghosting detector in Sect. 3. In Sect. 4, we present the subjective quality evaluation study of our image set. We compare the performance of our ghosting detector to the

subjective evaluation in Sect. 5, before the paper concludes in Sect. 6.

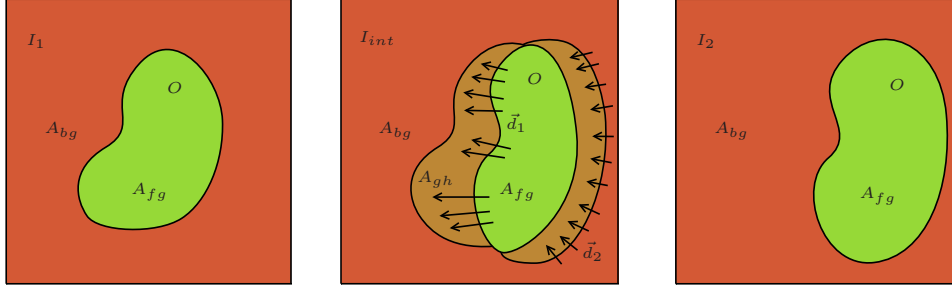


Figure 1: Ghosting occurs in image interpolation when the corresponding pixels from two or more original images I_1, I_2 are warped to different pixel positions in the interpolated image I_{int} : At the left border of O the region A_{fg} is replicated and displaced by \vec{d}_1 . The color values of A_{gh} appear alpha-blended from A_{fg} and A_{bg} . At the right border of O a similar displacement \vec{d}_2 occurs.

2 Related Work

Image quality assessment methods have been thoroughly studied over the past two decades. The main focus has been on the evaluation and improvement of image and video compression methods, such as JPEG2000, or MPEG-2.

Quality metrics have been developed to measure the quality of an image or video that has been altered due to compression or transmission over a communication channel. A good overview over existing metrics is provided by Engelke and Zepernik in [EZ07]. Basically, metrics can be separated into three classes: Full-reference (FR), Reduced-Reference (RR) and No-reference (NR) metrics.

Full-reference metrics compare a processed image to the original image [Mar04, WBSS04]. Distorted videos are evaluated in comparison to the original videos [WB01, SB07, SB05].

Reduced-reference metrics extract key features of the original image and compare them to key features extracted from the altered image. The key features are usually provided via an ancillary channel [WS05] or are embedded in the image [WWS⁺06].

No-reference metrics evaluate only the altered image and apply filters or heuristics to it. Liu et al. [LH08] detect the blockiness in compressed images and rely on the periodicity of artifacts due to the DCT transform. Sheikh et al. [SBC05] use natural scene statistics in the wavelet domain to assess the image quality to tackle the DWT coefficient quantization introduced by JPEG2000. A perceptual blur metric is introduced by Marziliano

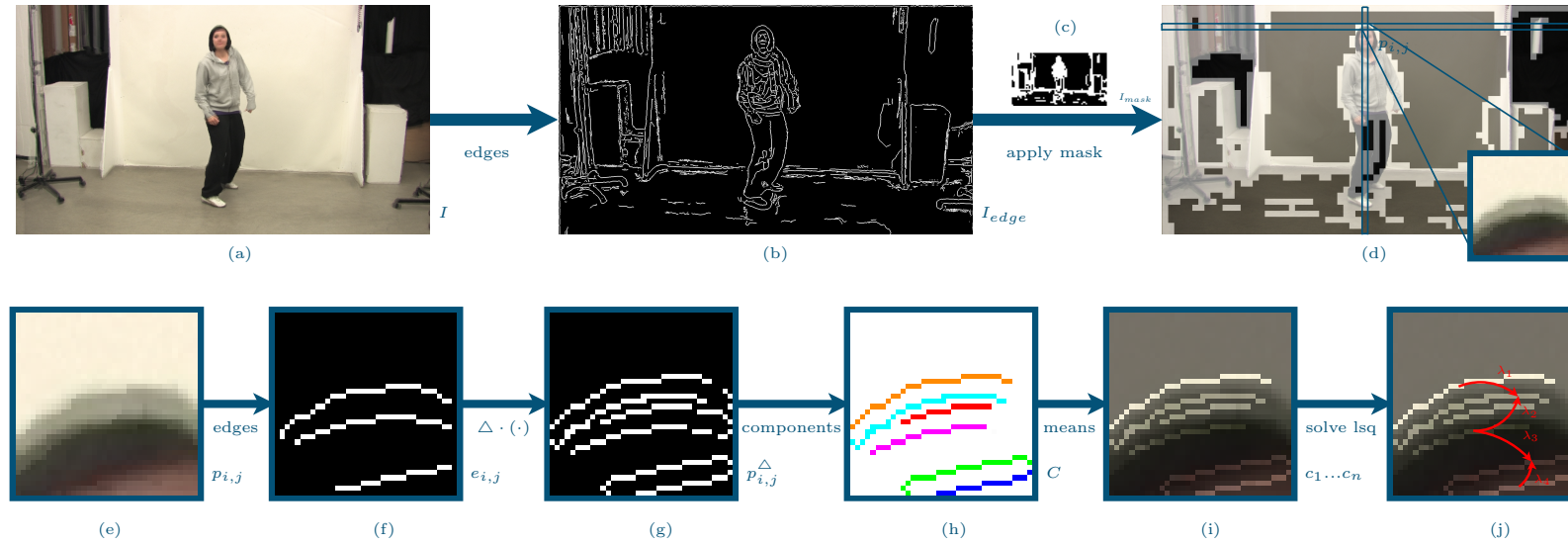


Figure 2: The processing steps of the algorithm. **Top row:** the input image I (a) is smoothed and an edge detection is applied to get I_{edge} (b). A mask (c) is calculated to obtain the relevant patches of I (d, bright regions). **Bottom row:** (from left to right): For each patch $p_{i,j}$ (e) the algorithm performs an edge detection to get $e_{i,j}$ (f) and the Laplacian operator $p_{i,j}^{\Delta}$ (g) is applied to it. Connected components in $p_{i,j}^{\Delta}$ are grouped into sets C (h). For each group $c \in C$ the mean color value of the corresponding image pixels in I is calculated (i). Finally a least-squares problem is solved for each three components in C (j). The red variables λ_1, λ_2 sum to 1, the 2-norm of the residual is small. The same holds for λ_3, λ_4 .

et al. [Mar04]. They measure the blurriness in terms of edge width, i.e. the pixel distance between local extrema around a detected edge pixel in gray-level images. Another blurriness metric searches for phase distortion in the Fourier spectrum[WS03]. Farias et al. [FM05] propose an artifact assessment in videos by combining a blockiness, blurriness and noisiness metric.

Only recently, quality measurements of free-viewpoint video results have been addressed. Kilner et al.[KSGH09, SKH08] investigate video errors caused by image rendering techniques and propose a reduced reference metric. They measured the pixel error of an image from a new viewpoint to the images of adjacent input cameras based on the Hausdorff-distance. However, their metric is only applicable in free-viewpoint systems based on geometric reconstruction. Our approach, in contrast, is purely image-based and does not rely of any geometric proxy. It examines only the interpolated image itself.

3 Ghosting artifact-detector

In order to assess image quality based on ghosting artifacts, we make two assumptions. We assume that ghosting can be detected locally, and that ghosting artifacts are only visible in areas containing strong object edges.

The details of the algorithm are explained further on in detail and are shown in Fig. 2.

In a first step, the input image I is subdivided into small patches $p_{i,j}$ of surface area d^2 . In our experiments, we chose $d = 15$ pixels. Since the most noticeable quality loss appears along object edges, the algorithm evaluates only patches near edges. According to our second assumption, object edges are usually predominant edges. In order to find them, an edge detection is applied to a low-pass filtered version of the input image I , retaining only the most prominent edges in I_{edge} , Fig. 2(b). The low-pass filtering is performed using a Gaussian kernel of size 10 with a standard deviation $\sigma = 10$. We use the Canny edge detector and set $\sigma = 1$ as standard deviation of its Gaussian filter. The low and high thresholds are chosen relative to the highest value of the gradient magnitude of the image.

In a second step (Fig. 2(c)), a binary mask I_{mask} is calculated from I_{edge} to determine the relevant patches. The mask image is d^2 times smaller than the input image I . A pixel $I_{i,j}$ in I_{mask} is set to 1, if its corresponding patch $p_{i,j}$ contains at least d edge pixels, Fig. 2(d). To account for edges traversing near patch boundaries, we dilate I_{edge} using a disk element with size 1.

After this preprocessing, the algorithm iterates over all selected patches P and assigns a label $l_{i,j} \in L$ with $L = \{ghosting, crisp\}$ to each patch $p_{i,j} \in P$.

In this assignment step the algorithm performs a Canny-edge detection on the patch $p_{i,j}$ of the input image I to obtain an edge patch $e_{i,j}$, Fig. 2(f).

To this edge patch the Laplacian operator is applied

$$p_{i,j}^{\Delta} = \Delta \cdot (e_{i,j}).$$

The Laplacian-transformed patch $p_{i,j}^{\Delta}$ contains nonzero pixels only in the 8-neighborhood of an edge pixel in $e_{i,j}$, Fig. 2(g). The main purpose of the Laplace operator is to perform a search for the adjacent pixels of an edge pixel. Alternatively, with more computational effort morphological methods can be employed. We assume that for each edge in $e_{i,j}$ there exist two connected components in $p_{i,j}^{\Delta}$, one connected component for each side of the edge. The algorithm then groups each connected component $c = \{I_{x_1,y_1}, \dots, I_{x_m,y_m}\}$ into a set $C = \{c_1, \dots, c_n\}$. Note that the color values of the input image are stored for each connected component, Fig. 2(h).

Afterwards (Fig. 2(i)) for each combination of any three connected components c_a, c_b, c_c , ($a \neq b \neq c$), the mean color values $\vec{m}_a, \vec{m}_b, \vec{m}_c \in \mathbb{R}^3$ are computed and a least-squares problem is solved:

$$(\vec{m}_a \ \vec{m}_b) \cdot \vec{\lambda} = \vec{m}_c, \vec{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, 0 \leq \lambda_1, \lambda_2 \leq 1$$

If there is a $\vec{\lambda}$ with $\lambda_1 + \lambda_2 = 1$ for three connected components c_a, c_b, c_c and with a small 2-norm of the residual, then the colors of c_a and c_b can be blended into the colors of c_c , Fig. 2(j). Therefore c_a, c_b, c_c are considered to belong to a ghosting artifact and the label $l_{i,j}$ of the patch is set as *ghosting*. This is due to the fact that ghosting occurs when a surface area is repeated, slightly shifted within the image, or blended with the background.

The magnitude of combinations can be restricted by different constraints, e.g. the size of each connected component's pixel area, the relative positions of the connected component's centroids, or a minimum color distance between any two connected components' color mean values.

The overall numerical quality $g(I)$ of the image is finally computed as the percentage of patches labeled as *ghosting* to the overall number of detected patches in the image:

$$g(I) = \frac{\sum_{p_{i,j} \in P} g(p_{i,j})}{\|P\|}, 0 \leq g(I) \leq 1$$

where

$$g(p_{i,j}) = \begin{cases} 1, & l_{i,j} = \textit{ghosting} \\ 0, & l_{i,j} = \textit{crisp} \end{cases}.$$



Figure 3: The three test scenes used for the subjective user study. From each scene original and interpolated images were evaluated.

4 User study evaluation

The presented metric has been evaluated against a subjective user study. For this study a set of 20 images per scene has been randomly assembled from 3 different scenes, Fig. 3, consisting of 50 % images from a video camera and 50 % rendered images from a novel viewpoint. The interpolated images have been generated with the algorithm proposed by Stich et al. [SLW⁺08a]. The rendered images from a novel viewpoint show quality alterations of different severity. In accordance with the international recommendations for subjective video quality assessment [ITU02], the image set was presented to 15 human observers in a laboratory environment. The observers were given the task to grade each of the presented images with one of the following values according to the ITU-R quality scale: "excellent" (highest score), "good", "fair", "poor", "bad" (lowest score).

In a second step, an online evaluation system was set up and advertised university-wide. Again the observers could grade the images with the same scores. The online evaluation was performed within the time span of a week, 61 students participated. In both studies the results have been analyzed as proposed by [ITU02]. For each dataset the mean opinion score and the 95% confidence interval per image are shown for the lab environment user study, Fig. 4, and the internet user study, Fig. 5. Comparing for each dataset the plot for the lab environment user study to the internet user study, it can be observed, that the confidence interval of the mean opinion score for most images in the lab user study is broader than in the internet evaluation. This can be explained by the small number of participants. Furthermore it can be stated for both user studies, that the mean opinion scores for interpolated images are lower than for original images. This statement holds for the skateboarder, firebreather, and dancer image set. The difference is smaller for the skateboarder since there are only few artifacts in the shadow region which seem to have less impact on perceived quality.

4 User study evaluation

7

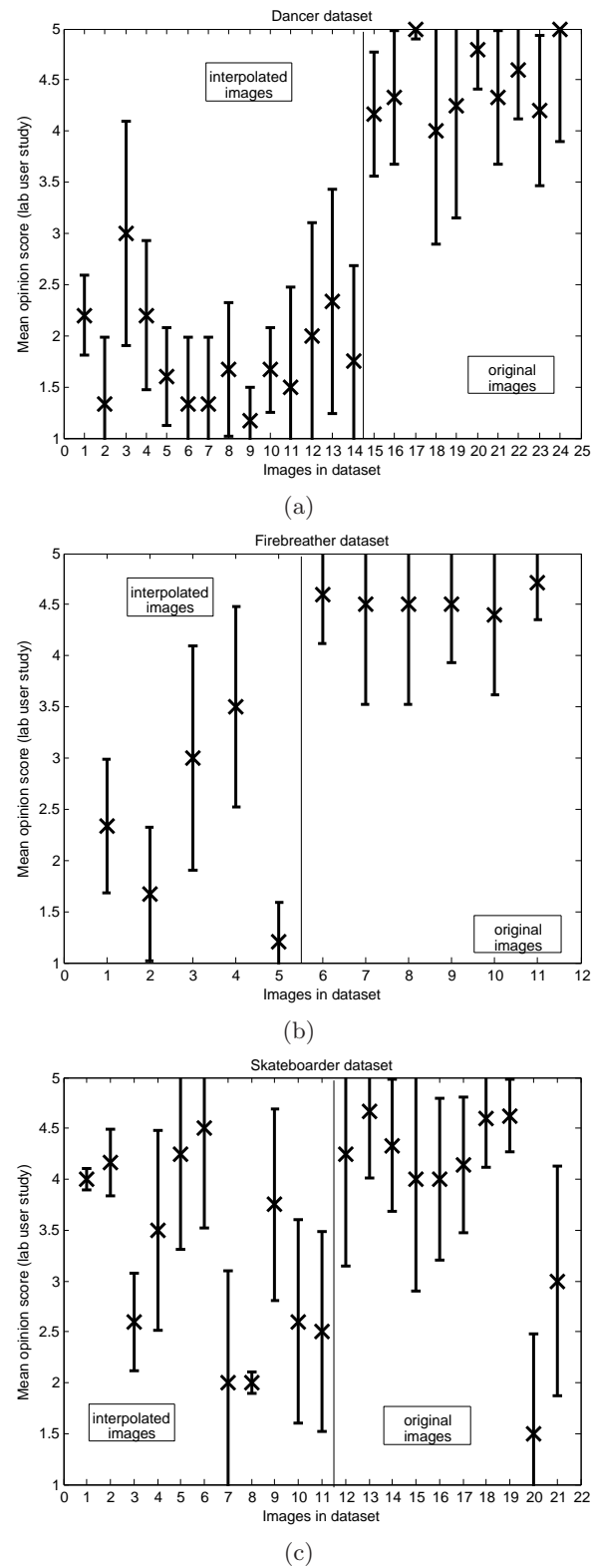
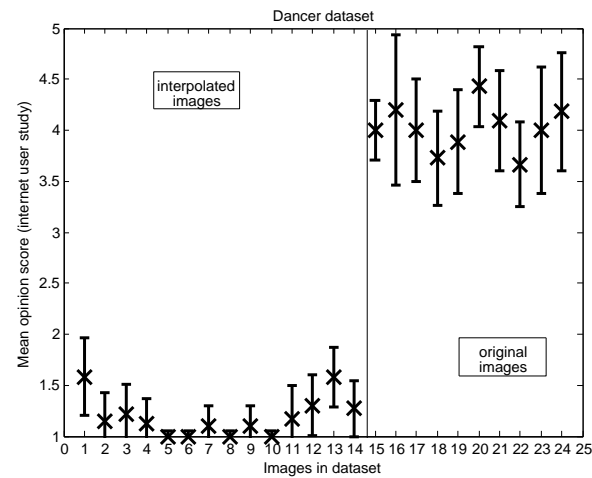


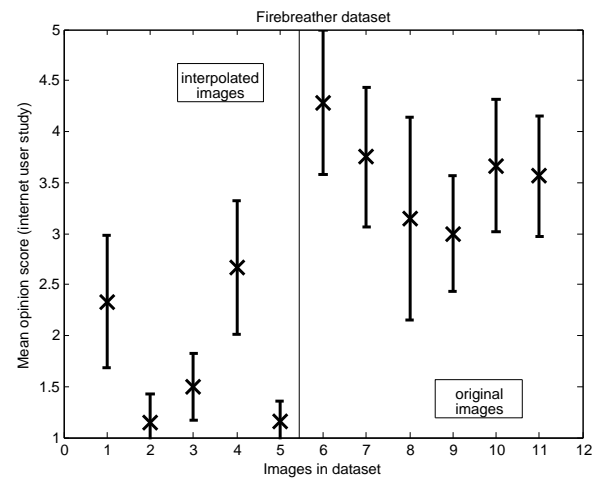
Figure 4: The results of the lab user study are plotted for each image dataset. Each plot shows the mean opinion score for a given image in the dataset. A vertical bar for each data point marks the 95% confidence interval. The vertical line distinguishes the interpolated images from the original images.

4 User study evaluation

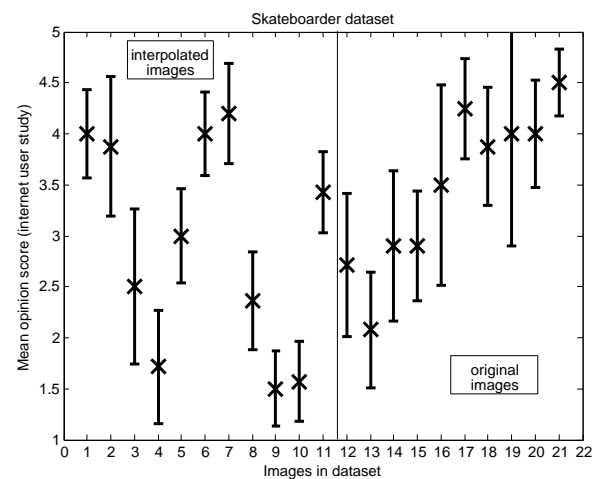
8



(a)



(b)



(c)

Figure 5: The results of the internet user studies are plotted for each image dataset. Each plot shows the mean opinion score for a given image in the dataset. A vertical bars for each data point marks the 95% confidence interval. The vertical line distinguishes the interpolated images from the original images.

5 Results

We have applied the algorithm described in Sect. 3 to the presented image dataset which consists of 3 different scenes with roughly 50% interpolated images and 50% original images each. The parameters of the algorithm have been set as described in Sect. 3 and remained unchanged for each scene. The image resolution of each image in the set was 960×540 pixels, the evaluation time varied between 10 and 30 seconds on a PC. In order to predict a subjective evaluation with the proposed metric, [ITU02] suggests that their result values have to be approximated by a logistic function $y = \frac{a \cdot b}{(a-b) \cdot \exp(-cx) + b}$, where x is the range of the metric, y is the mean opinion score and a, b, c are the parameters of the logistic function. In Fig. 6 such an approximation is depicted for each dataset. In each graph the results of the metric are approximated to the MOS values for each image in the dataset. A data point represents an image of the data set. The plots for the dancer, Fig. 6(a), the firebreather, Fig. 6(b), and the skateboarder, Fig. 6(c), dataset show a reasonable approximation by a logistic function. Hence the metric is capable of predicting a subjective value for an image in these datasets. From these results we deduce, that the proposed metric performs well on images which contain ghosting artifacts on the outline of opaque objects.

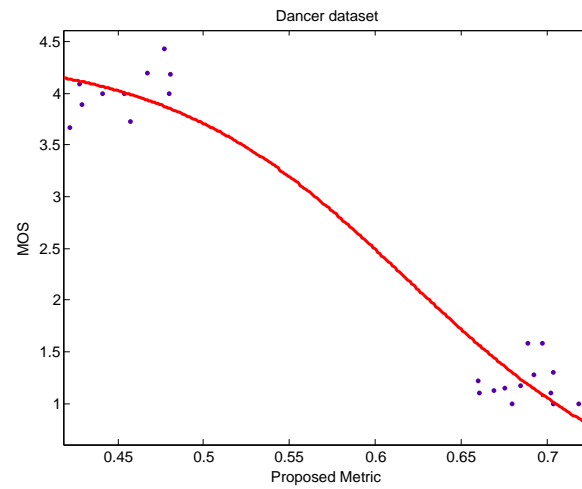
We also tested the proposed metric with two sets consisting of JPEG-compression and blurring artifacts. We applied a 90% JPEG-compression to the original images of the three scenes to get a set consisting only of blocky JPEG-artifacts. Furthermore we applied a Gaussian blur with kernel size 10 and $\sigma = 10$ to the original images of the three scenes to get a set consisting only of blurry regions. In the dataset consisting only of JPEG-compressed images, the metric labeled an average of 24.8% of the examined image patches falsely as *ghosting* (standard deviation $\sigma = 0.0805$). In the dataset consisting only of blurry images, the metric labeled an average of 10.8% of the examined image patches falsely as *ghosting* (standard deviation $\sigma = 0.0630$).

6 Conclusion

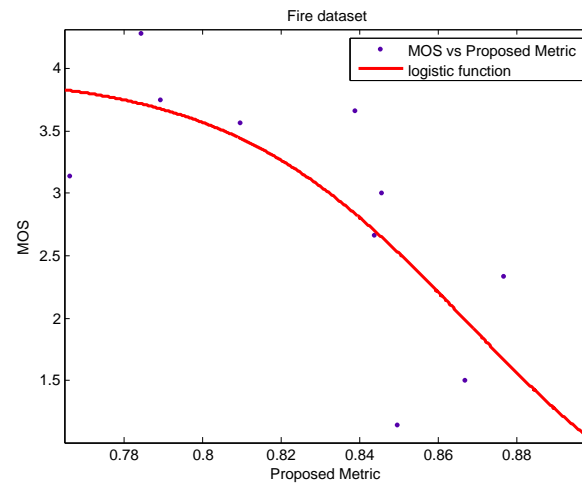
We have presented a versatile image quality metric which succeeds in detecting ghosting artifacts. The new metric is purely image-based and can be classified as no-reference metric. A laboratory study based on the international recommendations for subjective video quality assessment [ITU02] and a university-wide internet user study have been done to evaluate the confidence of the metric. Both studies are based on the same image dataset, which consist of 3 different scenes, roughly composed with 50% original and 50% interpolated images. Best accuracy is achieved for scenes with opaque objects and occlusion edges.

6 Conclusion

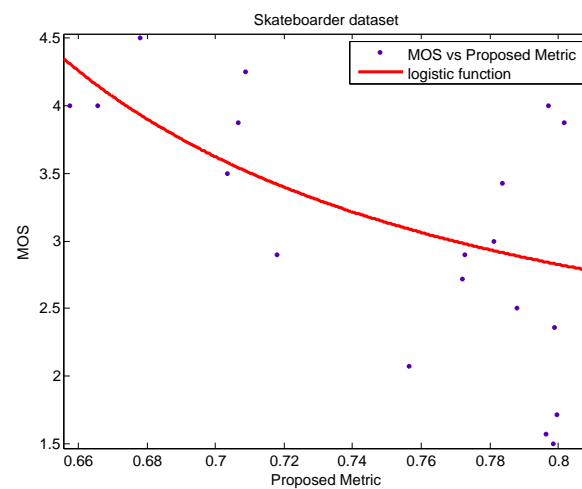
10



(a)



(b)



(c)

Figure 6: For each dataset the results of the proposed metric are compared to the results of the internet user study. For each image in the dataset the percentage of patches containing ghosting ($g(I)$) is plotted against the mean opinion score for the image. The red line indicates the approximation by a logistic function.

6 Conclusion

11

In the future, we want to extend our metric from still images to interpolated image sequences with a short time span to investigate the temporal evolution of ghosting artifacts. Secondly, we want to improve our metric towards non-opaque objects, like foam and water to get more accurate statements about their change of appearance during interpolation. In a third step we want to examine the correlation of the properties of a given patch to the overall perceived quality and introduce a weighting for each patch, e.g. according to the local contrast in the patch.

References

- [CTMS03] Joel Carranza, Christian Theobalt, Marcus A. Magnor, and Hans-Peter Seidel. Free-viewpoint video of human actors. *ACM Trans. on Graphics*, 22(3):569–577, July 2003.
- [dAST⁺08] Edilson de Aguiar, Carsten Stoll, Christian Theobalt, Naveed Ahmed, Hans-Peter Seidel, and Sebastian Thrun. Performance Capture from Sparse Multi-View Video. *ACM Trans. on Graphics*, 27(3):1–10, 2008.
- [EMD⁺08] Eisemann, M., De Decker, B., Magnor, M., Bekaert, P., De Aguiar, E., Ahmed, N., Theobalt, C., and A. Sellent. Floating textures. *Computer Graphics Forum*, 27(2):409–418, April 2008.
- [EZ07] U. Engelke and H. J. Zepernick. Perceptual-based quality metrics for image and video services: A survey. In *Next Generation Internet Networks, 3rd EuroNGI Conference on*, pages 190–197, 2007.
- [FM05] M. C. Q. Farias and S. K. Mitra. No-reference video quality metric based on artifact measurements. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III–141–4, 2005.
- [ITU02] R. ITU. Recommendation BT. 500-11. *methodology for the subjective assesment of the quality of televission pictures, tech. rep., International telecommunication union, Geneva, Switzerland*, 2002.
- [KSGH09] J. Kilner, J. Starck, J. Y. Guillemaut, and A. Hilton. Objective quality assessment in free-viewpoint video production. *Signal Processing: Image Communication*, 24(1-2):3–16, 2009.
- [LH96] M. Levoy and P. Hanrahan. Light Field Rendering. In *Proc. of ACM SIGGRAPH’96*, pages 31–42, New York, 1996. ACM Press/ACM SIGGRAPH.
- [LH08] Hantao Liu and I. Heynderickx. A no-reference perceptual blockiness metric. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 865–868, 2008.
- [Mar04] P. Marziliano. Perceptual blur and ringing metrics: application to jpeg2000. *Signal Processing: Image Communication*, 19(2):163–172, 2004.

REFERENCES

13

- [MP04] W. Matusik and H. Pfister. 3D TV: A Scalable System for Real-Time Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes. *ACM Trans. on Graphics*, 23(3):814–824, 2004.
- [SB05] H. R. Sheikh and A. C. Bovik. A visual information fidelity approach to video quality assessment. In *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 23–25, 2005.
- [SB07] Kalpana Seshadrinathan and Alan C. Bovik. An information theoretic video quality metric based on motion models. In *Proc. Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 25–26, 2007.
- [SBC05] H. R. Sheikh, A. C. Bovik, and L. Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *Image Processing, IEEE Transactions on*, 14(11):1918–1927, 2005.
- [SH07] J. Starck and A. Hilton. Surface Capture for Performance Based Animation. *IEEE Computer Graphics and Applications*, 27(3):21–31, 2007.
- [SKH08] J. Starck, J. Kilner, and A. Hilton. Objective quality assessment in free-viewpoint video production. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, pages 225–228, 2008.
- [SLW⁺08a] Timo Stich, Christian Linz, Christian Wallraven, Douglas Cunningham, and Marcus Magnor. Perception-motivated interpolation of image sequences. In *APGV '08: Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 97–106, New York, NY, USA, 2008. ACM.
- [SLW⁺08b] Timo Stich, Christian Linz, Christian Wallraven, Douglas Cunningham, and Marcus Magnor. Time and View Interpolation in Image Space. *Computer Graphics Forum (Proc. Pacific Graphics'08)*, 27(7):1781–1787, 10 2008.
- [VBK05] S. Vedula, S. Baker, and T. Kanade. Image Based Spatio-Temporal Modeling and View Interpolation of Dynamic Events. *ACM Trans. on Graphics*, 24(2):240–261, 2005.
- [WB01] Zhou Wang and Alan C. Bovik. A human visual system-based objective video. In *Intl Conf. On Multimedia Processing and Systems*, 2001.

REFERENCES

14

- [WBSS04] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.
- [WS03] Zhou Wang and Eero P. Simoncelli. Local phase coherence. In *in Adv. Neural Information Processing Systems (NIPS03*, volume 16, pages 786–792, 2003.
- [WS05] Zhou Wang and Eero P. Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *in Proc. of SPIE Human Vision and Electronic Imaging*, volume 5666, pages 149–159, 2005.
- [WSY07] H. Wang, M. Sun, and R. Yang. Space-Time Light Field Rendering. *IEEE Trans. Visualization and Computer Graphics*, pages 697–710, 2007.
- [WWS⁺06] Zhou Wang, Guixing Wu, H. R. Sheikh, E. P. Simoncelli, En-Hui Yang, and A. C. Bovik. Quality-aware images. *Image Processing, IEEE Transactions on*, 15(6):1680–1689, 2006.
- [ZKU⁺04] C. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-Quality Video View Interpolation Using a Layered Representation. *ACM Trans. on Graphics*, 23(3):600–608, 2004.